

WHAT IS CLAIMED IS:

1. A computer system comprising a processor and a memory, the processor being operable to initiate transactions involving the memory, the computer system further comprising a latency counter operable to generate a latency count for each of selected ones of the transactions, and a plurality of histogram counters, each histogram counter being operable to count selected ones of the latency counts corresponding to an associated latency range.

2. The computer system of claim 1 wherein the latency range associated with each histogram counter is programmable.

3. The computer system of claim 2 wherein each of the latency ranges is part of a window; a beginning value of the window being programmable.

4. The computer system of claim 1 wherein the latency counter is operable to count clock cycles between a first event and a second event associated with each of the selected transactions, the latency count for each of the selected transactions corresponding to a number of clock cycles.

5. The computer system of claim 4 wherein at least one of the first and second events is programmable.

6. The computer system of claim 5 wherein programming of the at least one of the first and second events results in the latency and histogram counters generating latency data for a specific transaction type.

5 7. The computer system of claim 1 wherein the latency counter is one of a plurality of latency counters, each latency counter being operable to generate the latency count for a portion of the selected transactions.

8. The computer system of claim 1 wherein the latency counter and the
10 histogram counters are operable to generate and count the latency counts at run-time.

9. The computer system of claim 8 wherein the processor is operable to alter a run-time parameter in response to latency information derived from the histogram counters.

15 10. The computer system of claim 1 wherein the processor is one of a plurality of processors operable to initiate the transactions.

11. The computer system of claim 10 wherein the processors and memory are interconnected with a point-to-point architecture.

20 12. The computer system of claim 10 wherein the processors and memory are interconnected with a shared-bus architecture.

13. The computer system of claim 10 wherein the processors are configured in a
25 plurality of processor clusters, each cluster including a plurality of local nodes and an

interconnection controller interconnected by a local point-to-point architecture, the interconnection controllers being operable to facilitate interaction among the clusters, and wherein the latency and histogram counters are implemented in each of the interconnection controllers.

5

14. The computer system of claim 13 wherein the interconnection controller in each cluster comprises a plurality of protocol engines for processing the transactions, and wherein at least one of the interconnection controller and the local nodes in each cluster is operable to map the transactions to the protocol engines according to destination information
10 associated with the transactions, and wherein the latency and histogram counters are implemented in each of the protocol engines.

15. The computer system of claim 14 wherein the plurality of protocol engines in each interconnection controller comprises at least one remote protocol engine for processing
15 first ones of the transactions targeting remote memory, and at least one local protocol engine for processing second ones of the transactions targeting local memory.

16. The computer system of claim 15 wherein the at least one remote protocol engine comprises a plurality of remote protocol engines, and the at least one local protocol
20 engine comprises a plurality of local protocol engines.

17. The computer system of claim 1 wherein the interconnection controllers are further operable to facilitate cache coherency across the computer system.

18. The computer system of claim 1 further comprising an input/output (I/O) device, wherein the processor is further operable to generate second transactions involving the I/O device, and wherein the latency counter is further operable to generate second latency counts for selected ones of the second transactions, and wherein the plurality of histogram counters are each operable to count selected ones of the second latency counts
5 corresponding to the associated latency range.

19. The computer system of claim 18 wherein the latency counter is operable to count clock cycles between a first event and a second event associated with each of the
10 selected second transactions, the latency count for each of the selected second transactions corresponding to a number of clock cycles.

20. The computer system of claim 19 wherein at least one of the first and second events is programmable.

21. The computer system of claim 20 wherein programming of the at least one of the first and second events results in the latency and histogram counters generating latency data for a specific transaction type.

22. An interconnection controller for use in a computer system having a plurality of processor clusters, each cluster including a plurality of local nodes and an instance of the interconnection controller interconnected by a local point-to-point architecture, the interconnection controller being operable to process transactions associated with the computer system, the interconnection controller further comprising a latency counter
25 operable to generate a latency count for each of selected ones of the transactions, and a

plurality of histogram counters, each histogram counter being operable to count selected ones of the latency counts corresponding to an associated latency range.

23. The interconnection controller of claim 22 wherein the latency range
5 associated with each histogram counter is programmable.

24. The interconnection controller of claim 23 wherein each of the latency ranges is part of a window, a beginning value of the window being programmable.

10 25. The interconnection controller of claim 22 wherein the latency counter is operable to count clock cycles between a first event and a second event associated with each of the selected transactions, the latency count corresponding to each of the selected transactions corresponding to a number of clock cycles.

15 26. The interconnection controller of claim 22 wherein the latency counter is one of a plurality of latency counters, each latency counter being operable to generate the latency count for a portion of the selected transactions.

20 27. The interconnection controller of claim 22 wherein the latency counter and the histogram counters are operable to generate and count the latency counts at run-time.

28. The interconnection controller of claim 27 wherein the interconnection controller is operable to alter a run-time parameter in response to latency information derived from the histogram counters.

29. The interconnection controller of claim 22 further comprising a plurality of protocol engines for processing the transactions, and circuitry which is operable to map the transactions to the protocol engines using destination information associated with the transactions, and wherein the latency and histogram counters are implemented in each of the
5 protocol engines.

30. An integrated circuit comprising the interconnection controller of claim 22.

31. The integrated circuit of claim 30 wherein the integrated circuit comprises an
10 application-specific integrated circuit.

32. At least one computer-readable medium having data structures stored therein representative of the interconnection controller of claim 22.

15 33. The at least one computer-readable medium of claim 32 wherein the data structures comprise a simulatable representation of the interconnection controller.

34. The at least one computer-readable medium of claim 33 wherein the simulatable representation comprises a netlist.
20

35. The at least one computer-readable medium of claim 32 wherein the data structures comprise a code description of the interconnection controller.

36. The at least one computer-readable medium of claim 35 wherein the code
25 description corresponds to a hardware description language.

37. A set of semiconductor processing masks representative of at least a portion of the interconnection controller of claim 22.

5 38. A computer-implemented method for measuring performance of a computer system, comprising:

generating a latency count for each of a plurality of transactions in the computer system; and

counting selected ones of the latency counts corresponding to each of a plurality of latency ranges, thereby generating latency distribution data.

39. The method of claim 38 further comprising dynamically altering a run-time parameter of the computer system in response to the latency distribution data.

15 40. The method of claim 39 wherein the run-time parameter relates to a data placement algorithm.

41. An electronic system characterized by a plurality of transactions, the electronic system comprising a latency counter operable to generate a latency count for each of selected ones of the transactions, and a plurality of histogram counters, each histogram counter being operable to count selected ones of the latency counts corresponding to an associated latency range.

42. The electronic system of claim 41 wherein the latency range associated with each histogram counter is programmable.

43. The electronic system of claim 42 wherein each of the latency ranges is part of a window, a beginning value of the window being programmable.

5 44. The electronic system of claim 41 wherein the latency counter is operable to generate the latency count between a first event and a second event associated with each of the selected transactions.

10 45. The electronic system of claim 44 wherein at least one of the first and second events is programmable.

15 46. The electronic system of claim 41 wherein the plurality of histogram counters are operable together to generate latency distribution data, the electronic system further comprising a processor which is operable to dynamically alter a run-time parameter of the electronic system in response to the latency distribution data.